NBS
RESEARCH
INFORMATION
CENTER

American National Standard

Adopted for Use by
the Federal Government

FIPS

FIPS PUB 90
See Notice on Inside
Front Cover

for information systems – optical character recognition (OCR) – guidelines for OCR print quality

ANSI X3.99-1983

JK
468
.A8A3
#90
1983

american national standards institute, inc
1430 broadway, new york, new york 10018

This guideline has been adopted for Federal Government use.

Details concerning its use within the Federal Government are contained in Federal Information Processing Standards Publication 90, Guideline for Optical Character Recognition (OCR) Print Quality. For a complete list of the publications available in the Federal Information Processing Standards Series, write to the Standards Processing Coordinator (ADP), Institute for Computer Sciences and Technology, National Bureau of Standards, Washington, D.C. 20234.

Remove this tab before mailing

Remove this tab before mailing

ANSI
X3.99-1983

American National Standard for Information Systems – Optical Character Recognition (OCR)– Guidelines for OCR Print Quality

Secretariat

Computer and Business Equipment Manufacturers Association

Approved July 1, 1983

American National Standards Institute, Inc

American National Standard

An American National Standard implies a consensus of those substantially concerned with its scope and provisions. An American National Standard is intended as a guide to aid the manufacturer, the consumer, and the general public. The existence of an American National Standard does not in any respect preclude anyone, whether he has approved the standard or not, from manufacturing, marketing, purchasing, or using products, processes, or procedures not conforming to the standard. American National Standards are subject to periodic review and users are cautioned to obtain the latest editions.

The American National Standards Institute does not develop standards and will in no circumstances give an interpretation of any American National Standard. Moreover, no person shall have the right or authority to issue an interpretation of an American National Standard in the name of the American National Standards Institute.

CAUTION NOTICE: This American National Standard may be revised or withdrawn at any time. The procedures of the American National Standards Institute require that action be taken to reaffirm, revise, or withdraw this standard no later than five years from the date of approval. Purchasers of American National Standards may receive current information on all standards by calling or writing the American National Standards Institute.

Published by

American National Standards Institute
1430 Broadway, New York, New York 10018

Copyright © 1983 by American National Standards Institute, Inc
All rights reserved.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

Printed in the United States of America

A5M983/6

Foreword

(This Foreword is not a part of ANSI X3.99-1983.)

These guidelines are supplemented by other OCR documents for character shapes, character positioning, inks, papers, and forms design.

The material contained in these guidelines originally appeared in American National Standard Character Set and Print Quality for Optical Character Recognition (OCR-A), ANSI X3.17-1977. In 1981, ANSI X3.17-1977 was revised and became a standard for OCR-A character shapes and size only (American National Standard Character Set for Optical Character Recognition (OCR-A), ANSI X3.17-1981). The print quality information that was deleted in the revision is now provided in these guidelines. No new quality requirements have been added.

This document, while based on OCR-A and OCR-B, can serve as a print quality guideline for other constant-stroke-width machine-readable fonts.

The terms, *uppercase* and *lowercase*, are widely used and accepted in the United States and have been used in these guidelines in place of the International Standards Organization terms, *letter* and *small letter*. In this document, these terms have the same meaning and can be used interchangeably.

Suggestions for improvement of this standard will be welcome. They should be sent to the Computer and Business Equipment Manufacturers Association, 311 First Street, NW, Suite 500, Washington, D.C. 20001.

This standard was processed and approved for submittal to ANSI by American National Standards Committee on Information Processing Systems, X3. Committee approval of the standard does not necessarily imply that all committee members voted for its approval. At the time it approved this standard, the X3 Committee had the following members:

Edward Lohse, Chair
Catherine A. Kachurik, Administrative Secretary

Organization Represented	*Name of Representative*
American Bankers Association	Cynthia Fuller Chris Crawford (Alt)
American Express Company	R. S. Newman R. G. Wilson (Alt)
American Library Association	Paul Peters
American Nuclear Society	Geraldine C. Main D. R. Vondy (Alt)
AMP Incorporated	Patrick E. Lannan C. Brill (Alt)
Association for Computing Machinery	J. A. N. Lee Pat Skelly (Alt)
Association of American Railroads	R. A. Petrash
Association of Computer Users	Hillel Segal Thomas Kurihara (Alt)
Association of the Institute for Certification of Computer Professionals	Thomas Kurihara Ardyn E. Dubnow (Alt)
Burroughs Corporation	Ira R. Purchis Jerrold S. Foley (Alt)
Control Data Corporation	Charles E. Cooper Keith Lucke (Alt)
Data General Corporation	Steven W. Weingart Howell A. Richards (Alt)
Data Processing Management Association	Ardyn E. Dubnow Joseph A. Federici (Alt)
Digital Equipment Computer Users Society	James Hodges John R. Barr (Alt)
Digital Equipment Corporation	Lois C. Frampton Gary S. Robinson (Alt)
Exxon Office Systems	Robert Greenblatt Richard Joyce (Alt)

Organization Represented	*Name of Representative*
General Services Administration	William C. Rinehuls
	Donald J. Page (Alt)
GUIDE International	Frank Kirshenbaum
	Leland Milligan (Alt)
Harris Corporation	Sam Mathan
	David Abmayr (Alt)
Hewlett-Packard	Donald C. Loughry
Honeywell Information Systems	Thomas J. McNamara
	David M. Taylor (Alt)
IBM Corporation	Mary Anne Gray
	J. S. Wilson (Alt)
IEEE Communications Society	Thomas A. Varetoni
IEEE Computer Society	Robert Poston
	Robert S. Stewart (Alt)
Lawrence Berkeley Laboratory	James A. Baker
	Robert J. Harvey (Alt)
Life Office Management Association	James J. Merrick
	James F. Foley, Jr (Alt)
Moore Business Forms, Inc	Delmer H. Oddy
National Bureau of Standards	Robert E. Rountree
	James H. Burrows (Alt)
National Communications System	Marshall L. Cain
	George W. White (Alt)
NCR Corporation	Thomas W. Kern
	William E. Snyder (Alt)
Perkin-Elmer	David Luckowicz
	James Pisarcik (Alt)
Prime Computer	Richard Kramer
	Winfried A. Burke (Alt)
Professional Secretaries International	Jerome Heitman
Recognition Technology Users Association	Herbert F. Schantz
	G. W. Wetzel (Alt)
SHARE, Inc	Thomas B. Steel
	Daniel Schuster (Alt)
Sperry Univac	Marvin W. Bass
	Charles D. Card (Alt)
Telephone Group	Henry L. Marchese
	Stuart H. Garland (Alt)
	Alvro E. Herron (Alt)
Texas Instruments, Inc	Presley Smith
	Richard Trow, Jr (Alt)
3M Company	R. C. Smith
Travelers Insurance Company, Inc	Joseph T. Brophy
U.S. Department of Defense	William LaPlant
	Belkis Leong-Hong (Alt)
Wang Laboratories, Inc	Carl W. Schwartz
	Marsha Hayek (Alt)
Xerox Corporation	John L. Wheeler
	Arthur R. Machell (Alt)

Subcommittee X3A1 on Optical Character Recognition, which developed these guidelines, had the following members:

T. C. Bagg, Chair
R. Monell, Vice-Chair
C. P. Newman, Secretary

C. Biss
R. Bloss
G. Brown
F. Cicha
J. L. Crawford
B. Daniels
W. Davidson
J. DeSautels
W. Foster
D. Gibson
K. Godwin
J. Hopkins
R. Ireland
T. Janning
C. Knoedel

G. Korzeniewski
J. McDonnell
S. McIntosh
R. Mestler
J. Miller
D. Newton
D. Oddy
T. Pealler
L. Richards
H. Schantz
N. Selke
E. Thompson
P. Traglia
N. Weiland
J. Wells

The X3A1 Working Group 3 on Print Quality, which had technical responsibility for the development of these guidelines, had the following members:

T. Bagg, Chair

K. Bye
J. Crawford
B. Frost
D. Gibson
S. Hou
R. Ireland
G. Martin
J. McDonnell
R. Mestler
J. Miller
D. Newton
T. Pealler
D. Rosenberg
E. Thompson
N. Weiland

Other persons who made technical contributions are:

R. Green
M. Hirsch
M. Hogan
A. Holt
H. Schantz
P. Traglia

Special thanks is given to Mrs. Candice Leatherman of the National Bureau of Standards for the preparation of the innumerable drafts for this manuscript.

Contents

American National Standard for Information Systems – Optical Character Recognition (OCR) – Guidelines for OCR Print Quality

1. Scope, Purpose, and Application

1.1 **Scope.** These guidelines describe the print quality parameters and measuring techniques for determining the quality of machine-printed characters to maximize the likelihood that they can be read by electro-optical means.

1.2 **Purpose.** The purpose of these guidelines is to establish procedures for determining suitability of machine-printed characters to be read optically.

1.3 **Application.** Because of the variety of OCR applications, these guidelines may not include all of the necessary print quality specifications or considerations for successful OCR reading. Items such as size of character set, document size, mechanical properties of the paper, and format details should be resolved by consultation between users, suppliers, and OCR manufacturers. Although each parameter may be independently specified, reading performance may deteriorate if the limits of two or more parameters are approached simultaneously. Every effort should be made to keep well within the limits. It is recognized, however, that in bulk printing from any device, the limits will occasionally and randomly be exceeded.

2. Related Standards

2.1 **American National Standards.** These guidelines are intended to be used in conjunction with the following American National Standards:

ANSI X3.17-1981, Character Set for Optical Character Recognition (OCR-A)

ANSI X3.49-1975 (R1982) Character Set for Optical Character Recognition (OCR-B)

ANSI X3.62-1979, Paper Used in Optical Character Recognition (OCR) Systems

ANSI X3.86-1980, Optical Character Recognition (OCR) Inks

ANSI X3.93M-1981, Optical Character Recognition (OCR) Character Positioning

2.2 **Other Publications.** These guidelines are intended to be used in conjunction with the following publications:

ASTM E 308-66 (1981), Standard Practice for Spectrophotometry and Description of Color in CIE 1931 System[1]

ASTM F 149-76 (1980), Standard Definitions of Terms Relating to Optical Character Recognition[1]

X3/TR-5-82, Design of Optical Character Recognition Forms (Technical Report)[2]

3. Definitions and Characteristics of the Printed Image

3.1 **General.** The performance of OCR systems depends to a large extent on the optical and dimensional properties of the printed image. For machine recognition of printed information, the paper shall provide high reflectance and printed images shall be sufficiently low reflectance to provide adequate contrast in the spectral range of interest.

All printed devices are not capable of producing images to the same tolerance ranges, and all readers are not capable of reliably recognizing characters in all tolerance ranges. In general, the performance level of an OCR reader will depend upon the number of characters to be recognized and the ability of the reader to tolerate a range of stroke widths, contrasts, voids, spots, and character position variations. Considering these factors, three ranges of print quality have been defined in 3.2. They describe the wide ranges of printing produced by various devices and processes and aid in

[1] Available from ASTM, 1916 Race Street, Philadelphia, PA 19103.

[2] Available from the Computer and Business Equipment Manufacturers Association, 311 First Street, NW, Suite 500, Washington, DC 20001.

classifying the capability of the associated reader.

It should be emphasized that reliable recognition of any specific character is related to the number of other characters in the character set to which the specific character must be compared. Thus, when using limited or partial character sets, such as in numeric-only applications, fully satisfactory performance can be achieved at a lower level of print quality. Equivalent performance levels using a larger character set generally require a higher level of print quality.

To achieve and maintain the required print quality for OCR, special precautions are necessary. These include judicious selection of supplies, more frequent adjustment and maintenance of printing equipment, and frequent changing of ribbons or checking of inks or other marking supplies.

This document contains definitions and information pertaining to the print quality characteristics of individual printed images without consideration of:

(1) Their position on the document

(2) Their spatial relationship to other character images

(3) Their relationship to any other nonOCR printing on the document

In many cases, the dimensional parameters can be determined visually using gages and magnified images. However, reflectance values, spectral response, etc, require special types of instruments to perform the measurements.

3.2 Print Quality Parameters and Tolerance Ranges. The quality of a printed image is established from a combination of measurable optical and dimensional parameters. Important print quality parameters include:

(1) Stroke-width variation

(2) Conformance to the character shape specified

(3) Reflectance within the character outline

(4) Reflectance outside the character outline

(5) Voids, spots, or extraneous ink areas

In any practical system, some tolerances must be permitted. The more restrictive the specified tolerance for each variable, the more likely it is that the print quality of the character will fall within the required range. To accommodate the variations in capability of various printing and reading devices, three ranges of print quality are recognized:

(1) *Range x*. Defines high-quality printing. This range places a greater burden on the printing or inking devices and a lesser burden on the reading device.

(2) *Range y*. Defines medium-quality printing. In an OCR system using this print quality range, the burden is more equally shared by the printing and the reading devices.

(3) *Range z*. Defines low-quality printing. This range places a lesser burden on the printing device and a greater burden on the reading device.

3.3 Definitions. OCR characters are designed to ultimately be read by machines. The scanning portion translates the optical and dimensional properties to electrical signals. The parameters in this section are defined in terms that are easily relatable to signals that a typical OCR scanner would produce while scanning the printing. The typical scanner is assumed to have a scanning aperture with a diameter of 0.2 mm (0.008 in). The limits given in these guidelines are based upon this size aperture. The limits for scanners with apertures of a different size must be extrapolated to determine the equivalent response.

The reflectance values are normalized to the document background reflectance (see discussion of print contrast signal [PCS], 3.3.1). Using the PCS ratio provides the advantage of allowing OCR readers to be self-calibrating. PCS can be converted to reflectance ratios if the background reflectance is known.

For additional definitions, see ASTM F 149-76 (1980).

3.3.1 Print Contrast Signal (PCS). Print contrast signal (PCS) is an expression that shows the relationship of the contrast between any selected point of interest within a character outline and the background. The PCS value of a point (p) is defined by the equation:

$$\mathrm{PCS} = \frac{R_w - R_p}{R_w}$$

where

R_w = the maximum reflectance found within the area of interest to which the PCS of point p is referenced. In measuring printed images, this area of interest shall be a rectangle approximately twice the nominal character centerline height (H) by twice the nominal character centerline width (W). The rectangle shall be centered vertically on the character being measured and the bottom shall be located $H/2$ below the character base line (see Figure 1).

R_p = the reflectance from a measured area centered on the point p.

The reflectance values R_w and R_p are measured within a circular area (aperture) having a diameter of 0.2 mm (0.008 in).

These reflectance specifications deal only with diffuse reflectance. Light that is specularly reflected shall be excluded. Reflectance values are normally referenced to barium sulfate or magnesium oxide. Absolute values are not necessary in the determination of PCS. The value of PCS depends only on the relative reflectance values of R_w and R_p.

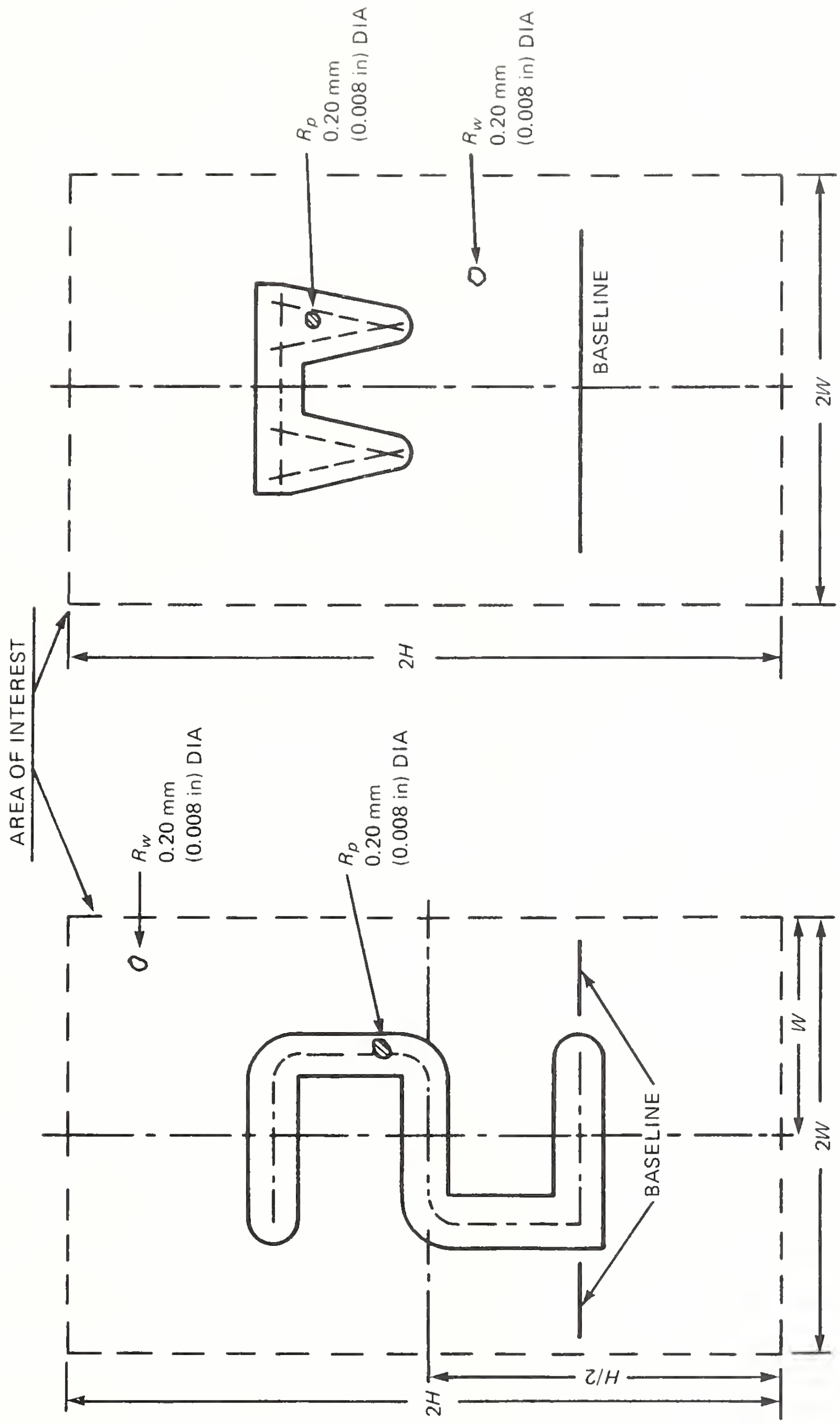


Figure 1
Illustration of Reflectance, R_w and R_p

Reflectance may be determined either by spectrophotometric measurements, or by a number of reflectance measurements in different spectral bands in accordance with ASTM E 308-66 (1981). This standard describes the equipment used for these measurements and the CIE color system.

The reflectance measurements shall be made using the infinite pad method (see 3.3.2) and in the spectral response range of the system that is to read the data (see 3.4).

3.3.2 Infinite Pad Method. The infinite pad method for measuring paper reflectance is the method in which the sample paper being measured is backed with enough thicknesses of the same type of paper so that doubling the number of sheets does not change the measured value of reflectance.

3.3.3 Character Outline Limit (COL). The dimensional properties of the printed image are measured by comparing the image with the character outline limit (COL) requirements. The COL is unique to each character in the set and to each specified stroke-width tolerance range. When constructed on a transparency for visual use, the COL is called a *COL gage.*

The COL requirements include definition of both minimum and maximum character outline limits. An example of a COL gage is shown in Figure 2. Except for the supplemental rules noted below, the COL is constructed, in all cases, from the nominal-stroke centerline graphic-shape outline drawing of the characters. The minimum or maximum COL may be considered as being generated by a circle, with a diameter equal to the minimum or maximum stroke width of the respective character size and tolerance range, following the character centerline.

The supplemental rules for modifying all circle-generated COL gages are as follows:

(1) The maximum COL at the free end of a stroke shall be squared off by ending the maximum stroke-width outline one-half the maximum stroke width beyond the end of the centerline. The ends of the outline shall be joined by lines perpendicular to those extended stroke outlines. The resultant corners are to be constructed as sharp as possible within the limits of the construction method. In no case are these corners to be rounded off with more than a radius of 0.10 mm (0.004 in).

(2) All internal corners on the maximum COL shall be rounded with a radius of 0.10 mm (0.004 in), except in the case of a character that has two internal corners closer than 0.20 mm (0.008 in), in which case the largest possible radius shall be used (for example, the numbers 1 and 8 and *z* stroke-width tolerance).

(3) When an OCR font shape standard specifies an external sharp corner, both the maximum and minimum COL outline shall be constructed with these corners sharp. The limits of COL construction methods may require some rounding of sharp corners. In no case shall these sharp corners be rounded off with more than a radius 0.10 mm (0.004 in).

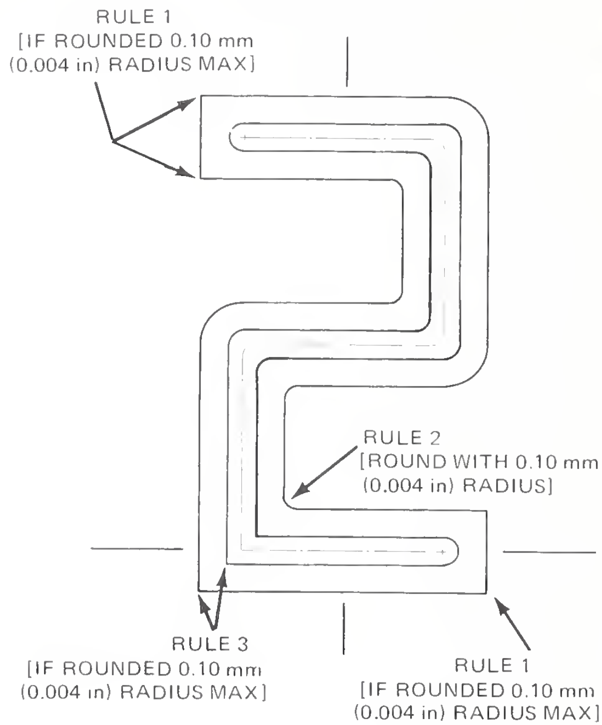


Figure 2
COL Gage

(4) Whenever any rounding of sharp corners due to COL gage construction limitations causes a portion of the character to violate the COL, the requirements of the font shape standard and these supplemental rules shall prevail.

It should be noted that having established the shape of the COL gage, it is not always possible to overlay the transparent COL gage on the exact printed character centerline because the centerline is not visible in the printed character. Because of this difficulty, the measured character is considered to be of the correct shape if the gage can be placed over the character in any position such that the edge requirements of Section 4 are satisfied.

3.3.4 Print Contrast Signal Degradation. Print contrast signal degradation is the result of poor inking with particular problems caused by the effects of voids and spots.

Voids are areas inside the minimum COL where the printing process did not deposit sufficient material on the background area and which have significantly higher

reflectance than other areas of the printed image. (For example, see Figure 3(a).)

Spots are areas outside the printed character with significantly lower reflectance than the surrounding background area. They are caused by dirt in the paper or smudging or splattering of ink during the printing process. (For examples, see Figure 3(a).)

Voids and spots are specified with two objectives in mind: (1) to allow physical measurement of the printed characteristics of a character and (2) to anticipate the scanner signal resulting from the actual scanning of the degraded character. (See Figures 3 and 4.)

The physical and scanning parameters of voids and spots shall be described as follows:

(1) The physical parameters are expressed in terms of size, shape, separation, and reflectance level, as measured by a scanning aperture of 0.2 mm (0.008 in). (See Figure 4.)

(2) The scanning parameters are expressed in terms of PCS_{void}, PCS_{spot}, PCS_{min}, and PCS_{max}, which are measures of the signal degradation related to voids and spots.

3.3.4.1 Physical Voids. Physical voids are defined as areas within the minimum COL, which can be identified as being lighter than a predetermined PCS level. This PCS level is designated the Void Threshold Level. Some voids are allowed, with the distinction between allowable and nonallowable voids being based on their size, separation, and PCS value (see Figure 4). All measurements are made with an aperture of 0.2 mm (0.008 in).

The criteria for the size, separation, and threshold level of physical voids are described in 3.3.4.1.1 through 3.3.4.1.3.

3.3.4.1.1 Void Size. Small voids shall be allowed if the distance (q) that the measuring aperture can be moved with a PCS equal to or less than the specified Void Threshold Level nowhere exceeds the distance given below in (1), (2), or (3). The movement shall be in a straight line and shall at all times remain inside the minimum COL. For any particular void, movement of the aperture shall be in only a single selected direction, i.e., a void may be somewhat oblong, but may exceed 0.2 mm (0.008 in) in only a single direction.

(1) *Range x Quality.* $q = 0$. No movement of aperture, i.e., only individual points are permitted.

(2) *Range y Quality.* $q \leqslant 0.1$ mm (0.004 in).

(3) *Range z Quality.* $q \leqslant 0.2$ mm (0.008 in).

3.3.4.1.2 Void Separation. The separation between any two allowable voids (that is where the PCS value is equal to or less than the specified Void Threshold Level) shall exceed 1.0 mm (0.040 in) center to center. Therefore, a small void that is allowable under the void size criteria shall become unallowable if it is within 1.0 mm (0.040 in) of any other void. (See Figure 4.)

3.3.4.1.3 Void Threshold Level. For instrumented measurement of physical voids, a PCS level shall be chosen such that an area within the COL with a measured PCS below this level is considered to be a candidate for being a void (subject to the size and separation criteria). This level should be consistent with the quality of printing desired and compatible with the capability of the OCR reader to be used to read the printing. A high threshold level is more restrictive of the printer while a lower threshold allows larger and lighter voids. (See Figure 4.)

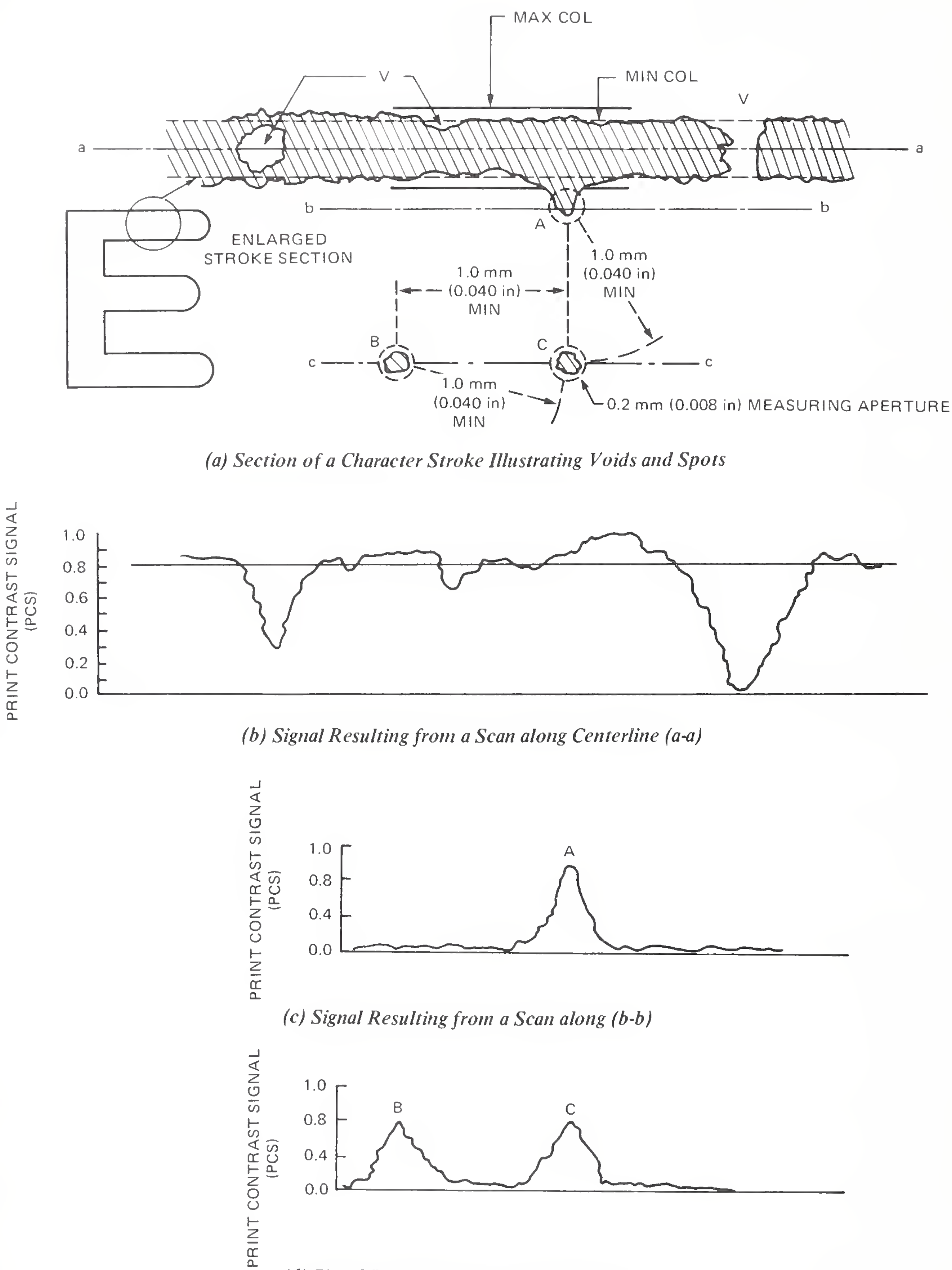
3.3.4.2 Physical Spots. Physical spots are defined as areas outside the maximum COL, but within the area of interest, which can be identified as being darker than a predetermined PCS level. This level is designated the Spot Threshold Level. Some spots are allowed, with the distinction between allowable and nonallowable spots being based on their size, separation, and PCS levels. Spots may be freestanding or may be connected to parts of the printed image. All measurements are made with an aperture of 0.2 mm (0.008 in). (See Figure 3 and Figure 4.)

The criteria for the size, separation, and threshold level of physical spots are described in 3.3.4.2.1 through 3.3.4.2.3.

3.3.4.2.1 Spot Size. The distance (q) that the PCS measuring aperture can be moved in a straight line with a PCS equal to or more than the Spot Threshold Level shall nowhere exceed the distances given in 3.3.4.1.1 for voids. The measuring aperture shall remain at all times outside the maximum COL. For any particular spot, movement of the aperture is permitted in only a single selected direction. (Thus, a spot may be somewhat oblong but may exceed 0.2 mm [0.008 in] in only a single direction.)

3.3.4.2.2 Spot Separation. The separation distance, center to center, between any two allowable spots (that is, where the PCS value is equal to or greater than the specified Spot Threshold Level but is within the size limits) shall exceed 1.0 mm (0.040 in) from the nearest spot.

3.3.4.2.3 Spot Threshold Level. For instrumented measurement of physical spots, a PCS level shall be chosen such that an area outside the COL with a measured PCS above this level is considered to be a candidate for being a spot (subject to the size and separation criteria). This level should be consistent with the quality of printing desired and the capability of the OCR reader to discriminate against spots. A lower threshold is more restrictive of the printer while a higher threshold level allows larger and darker spots.



(a) Section of a Character Stroke Illustrating Voids and Spots

(b) Signal Resulting from a Scan along Centerline (a-a)

(c) Signal Resulting from a Scan along (b-b)

(d) Signal Resulting from a Scan along (c-c)

Figure 3
Character Stroke Degradation and the Resulting Test Signal Device

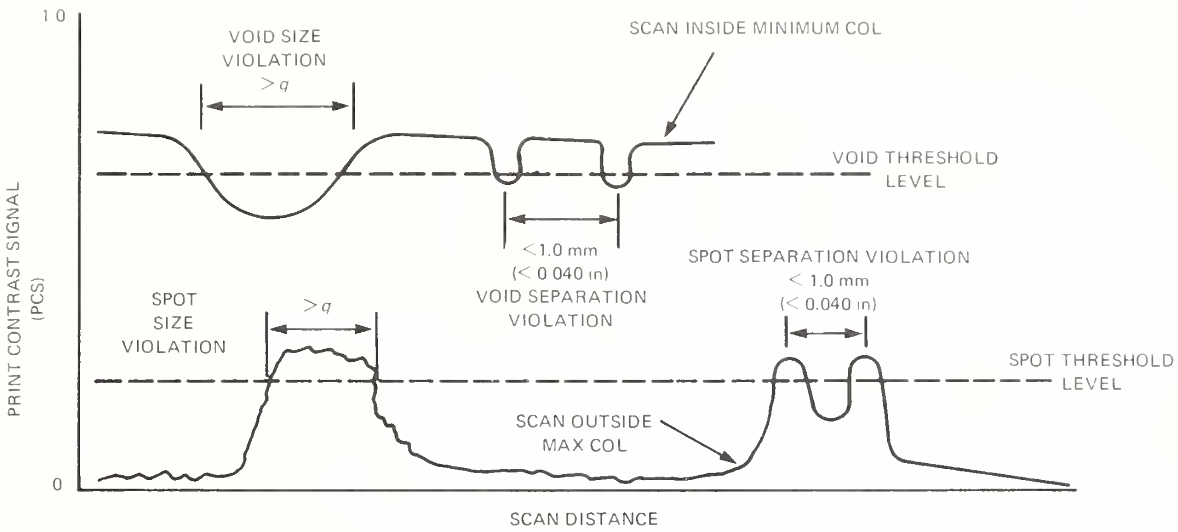


Figure 4
Physical Void and Spot Violation in Terms of Scan Signal

3.3.4.3 Scanning Signal Parameters – PCS_{void}/ PCS_{spot}

3.3.4.3.1 PCS_{void}. PCS_{void} is defined as the maximum Void Threshold Level at which all voids are allowable. For example, PCS_{void} represents the highest level to which the Void Threshold Level may be adjusted and at which no void will be detected using the criteria of 3.3.4.1, i.e., exceeds the size and separation limit for that range of quality (see Figure 5). It is a measure of the void free signal that is available to the optical reader.

3.3.4.3.2 PCS_{spot}. PCS_{spot} is defined as the minimum Spot Threshold Level at which all spots are allowable, i.e., PCS_{spot} represents the lowest level to which the Spot Threshold Level may be adjusted and at which no spot will be detected using the criteria of 3.3.4.2 (see Figure 6). It is a measure of the spot signal that the optical reader must overcome to be able to recognize characters reliably.

The ratio PCS_{void} and PCS_{spot} is a measure of the signal-to-noise ratio of the printed image. To reliably separate the character from the background noise, the ratio of PCS_{void}/PCS_{spot} should be high.

3.3.4.4 Contrast Variation Ratio. The variation of contrast within a character is expressed by the Contrast Variation Ratio (CVR), defined as:

$$CVR = \frac{PCS_{max}}{PCS_{min}}$$

where

PCS_{min} = the minimum PCS, which is the lowest PCS level that is not exceeded for a scanning distance of 0.20 mm (0.008 in) within the minimum COL. (See Figure 5.)

PCS_{max} = the maximum PCS, which is the highest PCS level that is continuously exceeded for a scanning distance of 0.20 mm (0.008 in) within the maximum COL. (See Figure 5.)

NOTE: For range z quality, PCS_{void} and PCS_{min} are the same by definition if PCS_{void} is determined by the void size criteria and not by the separation, i.e., the distance the aperture can be moved is 0.20 mm (0.008 in) in both cases.

3.4 Spectral Requirements for Measurement of PCS. Inks intended to be read by OCR systems require a high level of light absorption compared to the background, which is usually white paper. As a result, they produce a high PCS value ($PCS = (R_w - R_p)/R_w$). Carbon-pigmented inks absorb light over a broad spectral range. However, dye-based inks are more spectrally reflective and thus must be selected with care to produce high PCS values over ranges of interest. For the most part, dye-based inks are nonread inks for OCR systems.

The spectral response characteristics of OCR systems are dependent on the illumination source and the sensors. As a result, OCR reading devices are designed to

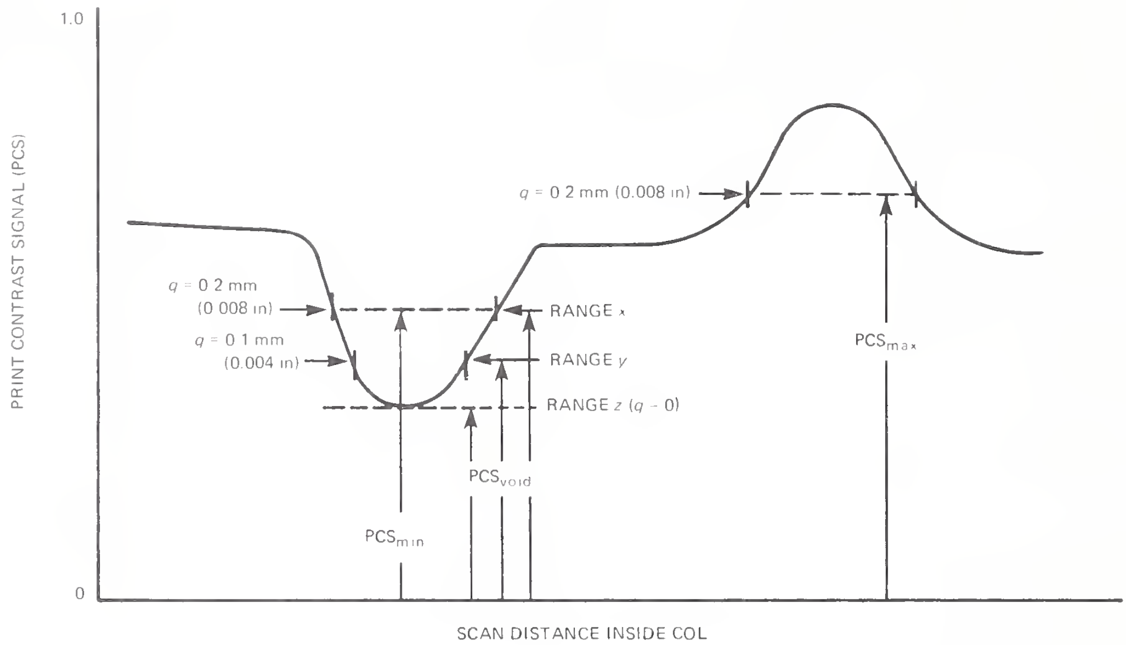


Figure 5
Measured PCS_{void}, PCS_{min}, and PCS_{max} for x, y, and z Ranges from the Same Stroke

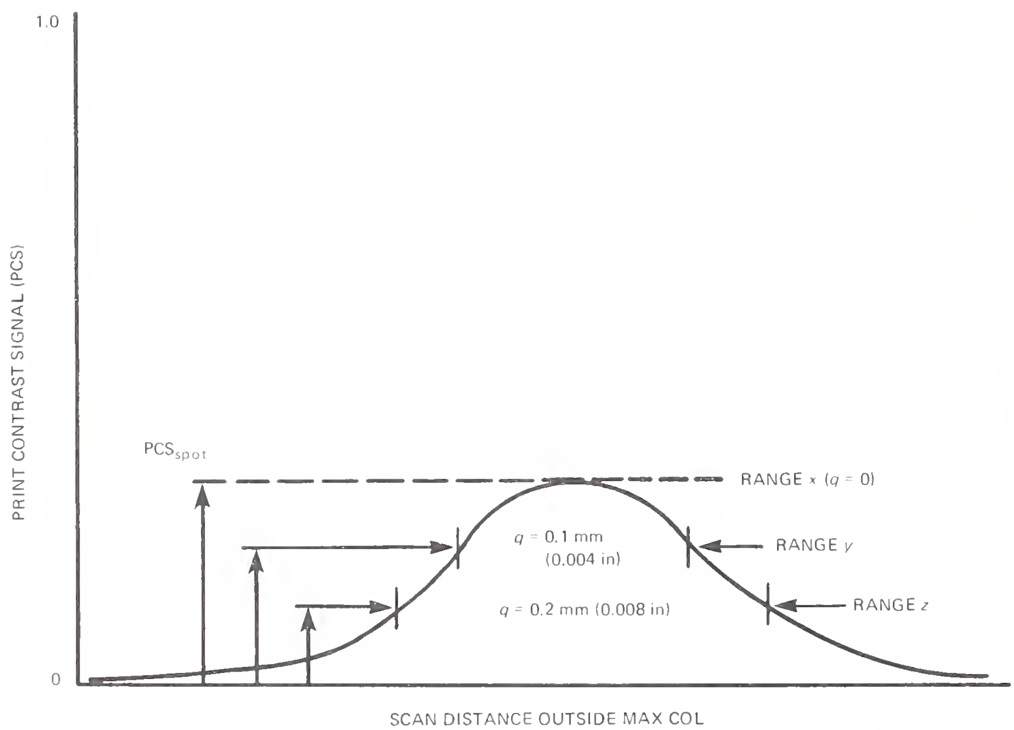


Figure 6
PCS_{spot} for x, y, and z Range Printing

Table 1
Nominal Character Sizes and Stroke-Width Tolerances

	Nominal Centerline				Normal Stroke Width (T)		Stroke-Width Tolerances*					
	Height (H)		Width (W)				Range x		Range y		Range z	
Size	mm	in	mm	in	mm	in	mm	in	mm	in	mm	in
I	2.40	0.094	1.40	0.055	0.36	0.014	±0.08	±0.003	±0.15	±0.006	±0.20	±0.008
III	3.20	0.126	1.52	0.060	0.38	0.015	±0.08	±0.003	±0.18	±0.007	+0.28 −0.18	+0.011 −0.007
IV	3.81	0.150	2.03	0.080	0.51	0.020	±0.13	±0.005	±0.25	±0.010	+0.41 −0.25	+0.016 −0.010

*Range tolerance does not apply to the Character Erase symbol.

NOTES:

(1) The height and width of OCR-B characters differ slightly from character to character. However, the largest OCR-B characters are no larger than the values shown; therefore, all requirements specified in this standard are applicable.

(2) When using lowercase letters of Size I in the OCR set, the range x tolerance should be applicable to all characters in the set.

(3) The lowercase letters i, j, m, p, and w of OCR-A were designed to exceed the nominal H or W values given in this table because of their unique characteristics.

(4) The dimensions in the metric and the inch-pound systems given in this standard are not precisely equivalent. For consistency, type designers may adopt either system but they should not be intermixed.

(5) Character Erase, Group Erase, Character Space, and Long Vertical Mark (LVM) were designed to exceed the nominal H or W values given in this table because of their unique characteristics.

(6) Size designation refers to ANSI OCR-A, as defined in ANSI X3.17-1981, and ANSI OCR-B, as defined in ANSI X3.49-1982. The Size II designation is reserved for international use.

have different spectral responses in the ultraviolet, the visible, or the near infrared regions. Carbon-pigmented inks should be considered as read inks on all OCR equipment. However, certain printing devices may not use carbon-pigmented inks and then readability may not be reliable in all spectral regions. Nonread inks can be formulated with a low absorption in the spectral range where the reader is most sensitive. At the same time, these OCR nonread inks can be very readable by humans.

The exact selection and formulation of these nonread inks should be determined by both the user and OCR equipment manufacturer for optimum results.

The spectral response of instruments to be used to measure and determine PCS should also be negotiated by the instrument manufacturer and the OCR equipment manufacturer to ensure compatibility. The following spectral bands have been used to measure PCS:

(1) Near ultraviolet

(2) Visual

(3) Infrared

See ANSI X3.86-1980 for a further description of OCR read and nonread inks.

4. Measurement of Character Geometry

4.1 General. The conformance of a printed image to its shape specification may be measured using COL gages. COL gages are designed to conform to the character set size as specified by the appropriate shape standard and tolerance range.

Table 1 shows the nominal centerline character height and width, the nominal stroke width for characters, and the stroke-width tolerance values for quality ranges x, y, and z. Each print quality range specifies the maximum permitted variation between the nominal stroke width and the actual stroke width of any character.

The COLs which are drawn using the values from Table 1 are normally displayed on a 1:1 scale for direct overlay on the printed character and virtually enlarged only by the magnification factor of a hand-held magnifier. For critical applications, 20:1 and even 50:1 COL enlargements are recommended to help determine character shape and edge irregularity conditions. COL gages do not define all the print quality parameters; they provide a simplified means for an approximate evaluation of the printed image shape quality for OCR characters used for data entry systems.

4.2 Determination of Stroke Edges. To determine if a printed image (character under test) is within specified tolerance limits, it is first necessary to establish the location of the stroke edge and then to determine if these stroke edges comply with the requirements relating to the character geometry. The location of the stroke edges of a character under test can be determined by instrumented measurements as follows or approximately as described in Section 6.

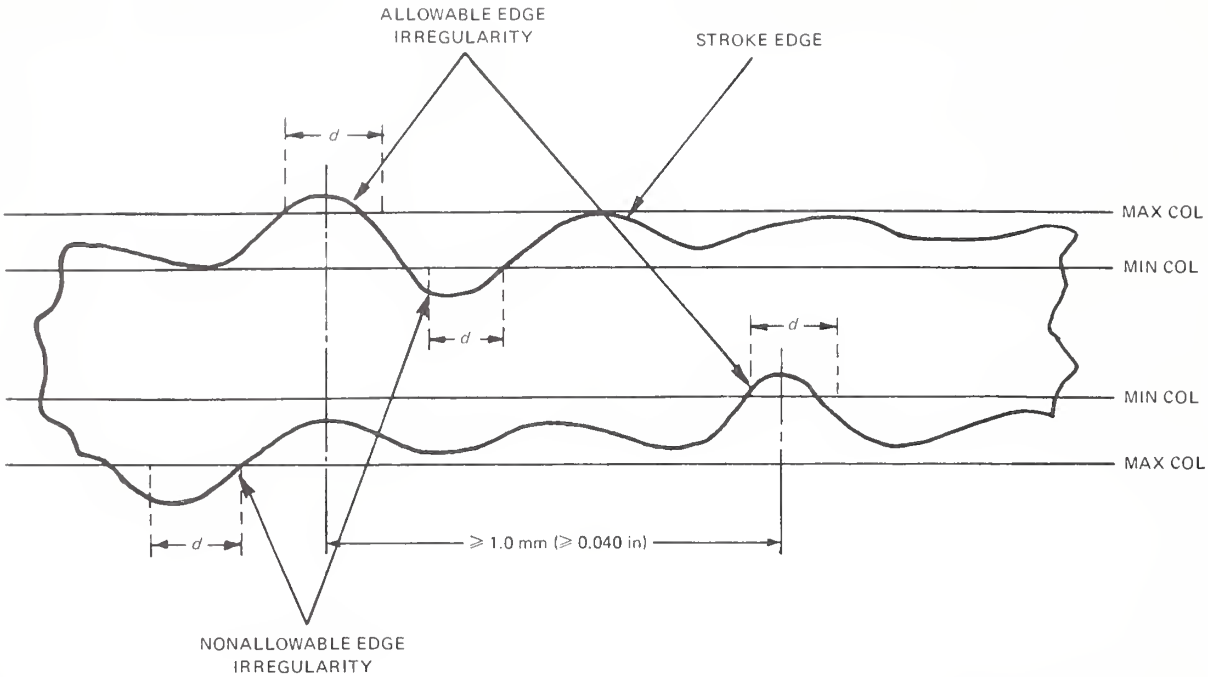


Figure 7
Enlarged View of Stroke Variation

Compliance with character geometry requirements depends on the ability to define the edge of the strokes. Stroke edges produced by high-quality printers (i.e., print quality range x) are relatively easy to identify visually. Instrumented methods should give results reasonably consistent with results observed visually. When stroke edges are not distinct due to smudges, nonuniform density, or shading at the edges, etc, the instrumented method may be necessary, and it should be closely related to the scanning and recognition systems to be used. Stroke edges are defined as the loci of points defined by the center of the scanning aperture when tracking around the character, using a threshold value of 1/2 PCS_{avg}. PCS_{avg} is the arithmetic average value of PCS measured over the darkest 80% of the character within the minimum COL.

4.3 Determination of Character Geometry for Print Quality Ranges. If any single position of the COL gage can be found where all strokes of the character defined by the stroke edges as determined above fill the minimum COL and, at the same time, do not extend beyond the boundaries of the maximum COL, then the character falls within the given print quality range. Small discrepancies, which are subject to further measurement as voids or spots, or both, should be ignored when making this measurement. However, under no circumstances, even if allowable under the definition of spots or voids given in Section 2, shall a small edge irregularity cross over the COL line for a distance exceeding the value d as measured along the COL line (see Figure 7). The value of d is:

Print Quality Range	*Distance (d)*	
	mm	*in*
x	0.30	0.012
y	0.40	0.016
z	0.50	0.020

The distance from any allowable edge irregularity to the nearest other allowable edge irregularity should be at least 1.0 mm (0.040 in) measured center to center along the COL line.

5. Measurement of Inking Quality

5.1 General. This section gives recommended reflectance values in terms of PCS or CVR of printed characters and describes methods for laboratory evaluation

of the optical characteristics that affect print quality.

The techniques for PCS and CVR determination require a significant number of measurements using sophisticated equipment, ranging from optical reflectometers to computer systems. This section describes measuring methods using such equipment. Section 6 describes simplified print quality evaluation methods that can be used in the field.

5.2 Contrast Variation Ratio (CVR) of a Character. For reliable reading, it is essential that the character area have a higher PCS value than the background. Reliability of reading may also be decreased as the unevenness or CVR of the ink density within the printed images is increased. For the tolerance ranges, the major portion of the character image within the minimum COL (normally interpreted as the measurement made along the character stroke virtual centerlines) should have PCS values greater and CVR values lower than shown in Table 2.

A higher PCS_{min} value, and conversely a lower CVR reading, result in improved OCR reader performance.

5.3 Voids. For the measurement of physical voids, it is recommended that the chosen Void Threshold Level be equal to or greater than 0.3 PCS. An area with a PCS_{void} equal to or less than 0.3 is a physical void.

5.4 Spots. It is recommended that the chosen Spot Threshold Level be less than 0.3 PCS and also less than 0.3 of the Void Threshold Level. To reliably separate the character from the background noise, the following limits of this ratio shall be observed:

$$\frac{PCS_{void}}{PCS_{spot}} \geqslant 1.5 \text{ for range } x$$

$$\frac{PCS_{void}}{PCS_{spot}} \geqslant 1.4 \text{ for range } y$$

$$\frac{PCS_{void}}{PCS_{spot}} \geqslant 1.3 \text{ for range } z$$

Table 2
PCS and CVR Values for Printed Characters

Range	Percentage of Measured PCS	Equal to or Greater Than	Remaining PCS_{min}	CVR
x	90%	0.70	≥0.47	≤1.50
y	85%	0.50	≥0.35	≤1.75
z	80%	0.35	≥0.30	≤2.00

NOTES:

(1) These values represent minimum requirements and every effort should be made to exceed them.

(2) Within a line or field of printed characters, PCS values should be as uniform as possible. This is of particular importance where more than one printing method is used to generate these characters.

6. Approximate Methods of Visually Evaluating Print Quality

6.1 General. In many cases, except for print contrast, OCR quality levels can be established with acceptable accuracy without having to employ sophisticated equipment. Stroke edges produced by high-quality printers (i.e., x tolerance range) are relatively easy to identify visually for most applications, and visual methods give results reasonably consistent with results obtained using reflectometers. When stroke edges are not distinct due to smudges, nonuniform density, or shading at the edges, etc, the instrumented method may be required and is recommended.

6.2 Visual Determination of Character Outline Limit. Frequently, a COL gage is used to determine visually if the printed character meets the outline requirements as indicated in the shape standard. Most commonly, the COL gage is merely a transparency depicting the specified maximum and minimum character outlines. In practice, the transparency is laid over the projected character image to determine its geometry relative to the specified shape. For critical applications, 20:1 and even 50:1 COL enlargements are used with a precison optical comparator to help determine character shape and edge irregularity conditions.

In constructing the COL gage character outlines, the instructions given in 3.2.2 should be followed.

It should be noted that having established the shape of the COL gage, it is not always possible to overlay the transparent COL gage on the exact printed character centerline because the centerline is not visible in the printed character. Because of this difficulty, the measured character is considered to be of the correct shape if the gage can be placed over the character in any position such that the edge requirements of Section 4 are satisfied.

6.3 Visual Identification of Allowable Voids. After the position for overlaying the COL gage on the character is determined,[3] a void is any area within the minimum COL that is devoid of ink.

[3] When measuring voids in characters that do not meet the stroke edge requirements, the COL gage is positioned to minimize the edge overlap of the COL.

A void is allowable if it satisfies the following conditons:

(1) It can be contained entirely within the areas defined below:

(a) *Range x*. A circle with a diameter of 0.2 mm (0.008 in), estimated visually.

(b) *Range y*. An area described by moving a circle with a diameter of 0.2 mm (0.008 in) along a straight line 0.1 mm (0.004 in) or less, estimated visually.

(c) *Range z*. An area described by moving a circle with a diameter of 0.2 mm (0.008 in) along a straight line 0.2 mm (0.008 in) or less, estimated visually.

(2) The separation between any two allowable voids that exceeds one-third the area of an inspection circle that is 0.2 mm (0.008 in) in diameter shall be at least 1.0 mm (0.040 in) (see Figure 8).

(3) Small voids or groups of small voids having a total area less than 1/3 the area of an inspection circle that is 0.2 mm (0.008 in) in diameter are allowed in unlimited number (see Figure 8).

6.4 Visual Identification of Allowable Spots. After the best fit between measuring gage and character is determined, any extraneous ink outside the maximum COL is defined as a spot. The sizes of spots determine if they are allowable or not. Any extraneous ink of sufficient area and nearly as dark as the lightest printing within the minimum COL may cause reading difficulties. Where there is a substantial question of subjective visual judgment, the instrumented method based upon PCS should be used.

A spot is allowable if it satisfies the following conditions:

(1) It can be contained entirely within the areas as defined below:

(a) *Range x*. A circle with a diameter of 0.2 mm (0.008 in), estimated visually.

(b) *Range y*. An area described by moving a circle with a diameter of 0.2 mm (0.008 in) along a straight line 0.1 mm (0.004 in) or less, estimated visually.

(c) *Range z*. An area described by moving a circle with a diameter of 0.2 mm (0.008 in) along a straight line 0.2 mm (0.008 in) or less, estimated visually.

(2) The separation between any two allowable spots that exceeds one-third the area of an inspection circle that is 0.2 mm (0.008 in) in diameter shall be at least 1.0 mm (0.040 in) (see Figure 9).

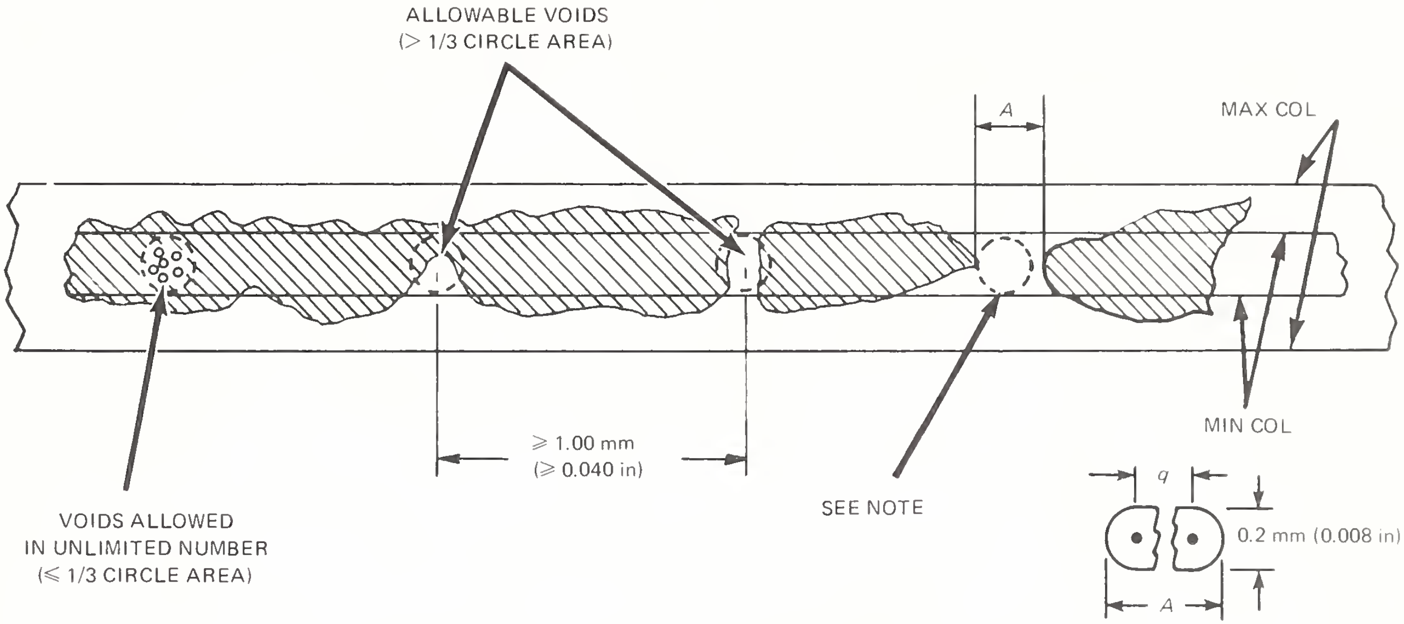
It is important to note one practical difference between the visual and PCS methods of assessing spots. With the visual method, the assessment depends only on the dimensions and spacing of the spots; it is independent of the density of the associated character. In PCS terms, when using the visual method of determining PCS, a spot associated with a dense character may be accepted while an identical spot associated with a weak character will be rejected.

The size and frequency of spots remote from a character should also be strictly controlled. Many reading operations are started upon detection of the first black point observed, and if a spot larger than 0.2 mm (0.008 in) occurs in any direction, the recognition process may begin. Spots larger than 0.2 mm (0.008 in) in any direction should be prevented.

6.5 Other Approximate Methods of Evaluation. Other less precise methods of visual measurement are sometimes used because they are portable and inexpensive.

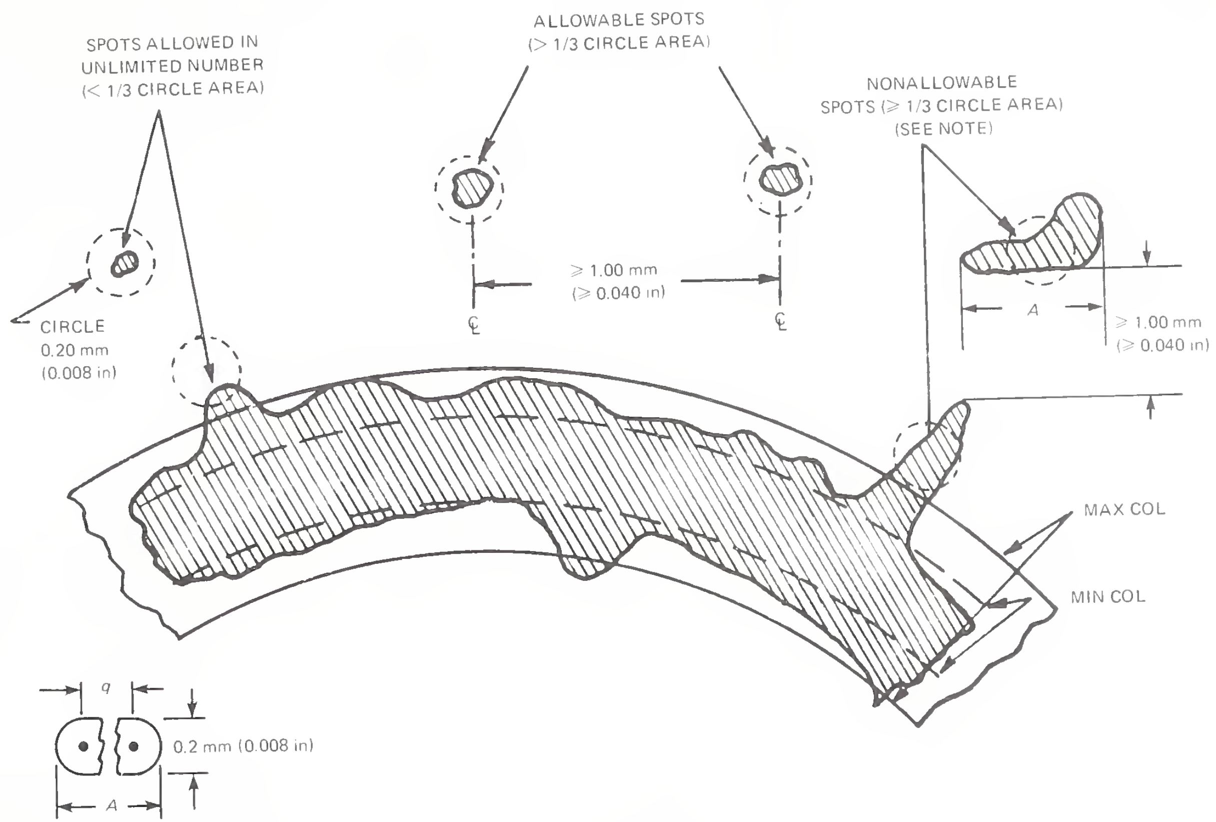
For some OCR fonts, small magnifiers (i.e., 5× to 10×) are employed. Representations of certain character outlines and various spot sizes are printed on a reticle placed in contact with the print.

Scales with characters reproduced at various levels of reflectance have also been used for certain applications. These allow for crude but frequently adequate visual evaluation of the image's PCS level.



NOTE: Nonallowable void occurs when q, the distance a 0.2-mm (0.008-in) circle can be moved within the void, exceeds:
(1) No movement range x, A = 0.2 mm (0.008 in)
(2) Half-circle diameter range y, A = 0.3 mm (0.012 in)
(3) Full-circle diameter range z, A = 0.4 mm (0.016 in)

Figure 8
Voids



NOTE: Nonallowable spot occurs when q, the distance a 0.2-mm (0.008-in) circle can be moved within the spot, exceeds:
(1) No movement range x, $A = 0.2$ mm (0.008 in)
(2) Half-circle diameter range y, $A = 0.3$ mm (0.012 in)
(3) Full-circle diameter range z, $A = 0.4$ mm (0.016 in)

Figure 9
Spots

AMERICAN NATIONAL STANDARD NO. ____________

PLEASE FILL IN ANSI DESIGNATION

Only One Number Per Card

To ANSI Sales Department

Please advise me when new editions of this standard are issued.

PLEASE PRINT

NAME ____

COMPANY ____

ADDRESS ____

____

City ______________________ State ________ Zip ________

FIRST CLASS MAIL

Place
Stamp
Here

American National Standards Institute
1430 Broadway
New York, N.Y. 10018

Attention: Sales Department

American National Standards for Information Processing Systems

X3.1-1976 Synchronous Signaling Rates for Data Transmission
X3.2-1970 (R1976) Print Specifications for Magnetic Ink Character Recognition
X3.3-1970 (R1976) Bank Check Specifications for Magnetic Ink Character Recognition
X3.4-1977 Code for Information Interchange
X3.5-1970 Flowchart Symbols and Their Usage
X3.6-1965 (R1973) Perforated Tape Code
X3.9-1978 Programming Language FORTRAN
X3.11-1969 General Purpose Paper Cards
X3.14-1973 Recorded Magnetic Tape (200 CPI, NRZI)
X3.15-1976 Bit Sequencing of the American National Standard Code for Information Interchange in Serial-by-Bit Data Transmission
X3.16-1976 Character Structure and Character Parity Sense for Serial-by-Bit Data Communication in the American National Standard Code for Information Interchange
X3.17-1981 Character Set for Optical Character Recognition (OCR-A)
X3.18-1974 One-Inch Perforated Paper Tape
X3.19-1974 Eleven-Sixteenths-Inch Perforated Paper Tape
X3.20-1967 (R1974) Take-Up Reels for One-Inch Perforated Tape
X3.21-1967 Rectangular Holes in Twelve-Row Punched Cards
X3.22-1973 Recorded Magnetic Tape (800 CPI, NRZI)
X3.23-1974 Programming Language COBOL
X3.24-1968 Signal Quality at Interface between Data Processing Terminal Equipment and Synchronous Data Communication Equipment for Serial Data Transmission
X3.25-1976 Character Structure and Character Parity Sense for Parallel-by-Bit Data Communication in the American National Standard Code for Information Interchange
X3.26-1980 Hollerith Punched Card Code
X3.27-1978 Magnetic Tape Labels and File Structure
X3.28-1976 Procedures for the Use of the Communication Control Characters of American National Standard Code for Information Interchange in Specified Data Communication Links
X3.29-1971 Specifications for Properties of Unpunched Oiled Paper Perforator Tape
X3.30-1971 Representation for Calendar Date and Ordinal Date
X3.31-1973 Structure for the Identification of the Counties of the United States
X3.32-1973 Graphic Representation of the Control Characters of American National Standard Code for Information Interchange
X3.34-1972 Interchange Rolls of Perforated Tape
X3.36-1975 Synchronous High-Speed Data Signaling Rates between Data Terminal Equipment and Data Communication Equipment
X3.37-1980 Programming Language APT
X3.38-1972 (R1977) Identification of States of the United States (Including the District of Columbia)
X3.39-1973 Recorded Magnetic Tape (1600 CPI, PE)
X3.40-1983 Unrecorded Magnetic Tape (9-Track 800 CPI, NRZI; 1600 CPI, PE; and 6250 CPI, GCR)
X3.41-1974 Code Extension Techniques for Use with the 7-Bit Coded Character Set of American National Standard Code for Information Interchange
X3.42-1975 Representation of Numeric Values in Character Strings
X3.43-1977 Representations of Local Time of the Day
X3.44-1974 Determination of the Performance of Data Communication Systems
X3.45-1982 Character Set for Handprinting
X3.46-1974 Unrecorded Magnetic Six-Disk Pack (General, Physical, and Magnetic Characteristics)
X3.47-1977 Structure for the Identification of Named Populated Places and Related Entities of the States of the United States for Information Interchange
X3.48-1977 Magnetic Tape Cassettes (3.810-mm [0.150-Inch] Tape at 32 bpmm [800 bpi], PE)
X3.49-1975 Character Set for Optical Character Recognition (OCR-B)
X3.50-1976 Representations for U.S. Customary, SI, and Other Units to Be Used in Systems with Limited Character Sets
X3.51-1975 Representations of Universal Time, Local Time Differentials, and United States Time Zone References
X3.52-1976 Unrecorded Single-Disk Cartridge (Front Loading, 2200 BPI) (General, Physical, and Magnetic Requirements)
X3.53-1976 Programming Language PL/I
X3.54-1976 Recorded Magnetic Tape (6250 CPI, Group Coded Recording)
X3.55-1982 Unrecorded Magnetic Tape Cartridge, 0.250 Inch (6.30 mm), 1600 bpi (63 bpmm), Phase encoded
X3.56-1977 Recorded Magnetic Tape Cartridge, 4 Track, 0.250 Inch (6.30 mm), 1600 bpi (63 bpmm), Phase Encoded
X3.57-1977 Structure for Formatting Message Headings Using the American National Standard Code for Information Interchange for Data Communication Systems Control
X3.58-1977 Unrecorded Eleven-Disk Pack (General, Physical, and Magnetic Requirements)
X3.59-1981 Magnetic Tape Cassettes, Dual Track Complementary Return-to-Bias (CRB) Four-States Recording on 3.81-mm (0.150-Inch) Tape
X3.60-1978 Programming Language Minimal BASIC
X3.61-1978 Representation of Geographic Point Locations
X3.62-1979 Paper Used in Optical Character Recognition (OCR) Systems
X3.63-1981 Unrecorded Twelve-Disk Pack (100 Megabytes) (General, Physical, and Magnetic Requirements)
X3.64-1979 Additional Controls for Use with American National Standard Code for Information Interchange
X3.66-1979 Advanced Data Communication Control Procedures (ADCCP)
X3.72-1981 Parallel Recorded Magnetic Tape Cartridge, 4 Track, 0.250 Inch (6.30 mm), 1600 bpi (63 bpmm), Phase Encoded
X3.73-1980 Single-Sided Unformatted Flexible Disk Cartridge (for 6631-BPR Use)
X3.74-1981 Programming Language PL/I, General-Purpose Subset
X3.76-1981 Unformatted Single-Disk Cartridge (Top Loading, 200 tpi 4400 bpi) (General, Physical, and Magnetic Requirements)
X3.77-1980 Representation of Pocket Select Characters
X3.79-1981 Determination of Performance of Data Communications Systems That Use Bit-Oriented Communication Procedures
X3.80-1981 Interfaces between Flexible Disk Cartridge Drives and Their Host Controllers
X3.82-1980 One-Sided Single-Density Unformatted 5.25-Inch Flexible Disk Cartridge (for 3979-BPR Use)
X3.83-1980 ANSI Sponsorship Procedures for ISO Registration According to ISO 2375
X3.84-1981 Unformatted Twelve-Disk Pack (200 Megabytes)(General, Physical, and Magnetic Requirements)
X3.85-1981 1/2-Inch Magnetic Tape Interchange Using a Self-Loading Cartridge
X3.86-1980 Optical Character Recognition (OCR) Inks
X3.88-1981 Computer Program Abstracts
X3.89-1981 Unrecorded Single-Disk, Double-Density Cartridge (Front Loading, 2200 bpi, 200 tpi) (General, Physical, and Magnetic Requirements)
X3.91M-1982 Storage Module Interfaces
X3.92-1981 Data Encryption Algorithm
X3.93M-1981 OCR Character Positioning
X3.95-1982 Microprocessors — Hexadecimal Input/Output, Using 5-Bit and 7-Bit Teleprinters
X3.96-1983 Continuous Business Forms (Single-Part)
X3.98-1983 Text Information Interchange in Page Image Format (PIF)
X3.99-1983 Print Quality Guideline for Optical Character Recognition (OCR)
X3.102-1983 Data Communication Systems and Services — User-Oriented Performance Parameters
X3.103-1983 Unrecorded Magnetic Tape Minicassette for Information Interchange, Coplanar 3.81 mm (0.150 in)
X3.104-1983 Recorded Magnetic Tape Minicassette for Information Interchange, Coplanar 3.81 mm (0.150 in), Phase Encoded
X3.105-1983 Data Link Encryption
X3.106-1983 Modes of Operation for the Data Encryption Algorithm
X11.1-1977 Programming Language MUMPS
IEEE 416-1978 Abbreviated Test Language for All Systems (ATLAS)
IEEE 716-1982 Standard C/ATLAS Language
IEEE 717-1982 Standard C/ATLAS Syntax
IEEE 770X3.97-1983 Programming Language PASCAL
IEEE 771-1980 Guide to the Use of ATLAS
MIL-STD-1815A-1983 Reference Manual for the Ada Programming Language

---

X3/TR1-82 Dictionary for Information Processing Systems (Technical Report)

August 1983